DA6233

Kilger

Exercise #4 Correspondence Analysis

Name\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. You are analyzing the relationship between race and pizza for women for a market research firm. You can find the frequencies of this variables in the pdf file in the exercise 4 folder. You can also find the data file in that folder as well in csv format. You can suck this into SAS, SPSS, R, Python or anything else except Excel – I hate Excel and it is not a data analytics platform. This exercise does however work best under SAS.
2. Run a correspondence analysis on the data. Then answer the following questions.
   1. Show the contingency table. Examining the table, is it easy to see relationships between race categories and pizza restaurants using this table?
   2. Generate and show a table that shows the expected frequencies for each cell.
   3. Generate and show a table that shows the chi square contribution to the total chi square for the table for each cell in the contingency table. Which cell contributes the most to the total chi square? Which cell contributes the least to the total chi square for the table?
   4. Using the total chi square for the table, calculate the degrees of freedom for the total chi square, find the critical value in the chi square table for those degrees of freedom and state the total chi square for the table from part c above. Can you reject the null hypothesis that there is no relationship between race and pizza restaurant?
   5. Generate and show row and column tables for mass, inertia and quality. Which row has the highest inertia? Which column has the highest quality?
   6. Generate and show the correspondence map for this analysis. What proportion of the variance does dimension 1 explain? What proportion of the variance does dimension explain?
   7. Find an example in the correspondence plot where the angle between row and column values suggests a strong association – draw the lines from the origin to these two points on the plot.
   8. Generate and show the standardized adjusted residuals for each of the cells in the contingency table. Which of these cells passes the test suggested by Sharpe for contributing significantly to the total chi square?

**Be sure to cut and paste your tables into your report along with your interpretation. Also be sure to include a copy of your output from the statistical package.**